American Journal of
EPIDEMIOLOGY

Volume 154
Number 5
September 1, 2001

Sponsored by the Society for Epidemiologic Research
Published by Oxford University Press

COMMENTARY

Using Validation Sets for Outcomes and Exposure to Infection in Vaccine Field Studies

M. Elizabeth Halloran and Ira M. Longini, Jr.

Methods of adjusting for bias in estimates due to mismeasured or missing covariates and outcomes through the use of validation sets have been developed in many types of health studies. These methods can be employed for the efficient design and analysis of vaccine studies as well. On the one hand, nonspecific case definitions can lead to attenuated efficacy and effectiveness estimates, but confirmation by culture or a quick test of the infectious agent is also expensive and difficult. On the other hand, data on exposure to infection can influence estimates of vaccine efficacy, but good data on exposure are difficult to obtain. In this paper, the authors show how use of small validation sets can correct the bias of the estimates obtained from a large main study while maintaining efficiency. They illustrate the approach for outcomes using the example of influenza vaccine efficacy and effectiveness trials and illustrate the approach for exposure to infection using the example of a human immunodeficiency virus vaccine trial. The authors discuss challenges posed by infectious diseases in the use of currently available methods. Development of these efficient designs and methods of analysis for vaccine field studies will improve estimation of vaccine efficacy for both susceptibility and infectiousness, as well as estimation of indirect and overall effects of vaccination in community trials. *Am J Epidemiol* 2001;154: 391–8.

bias (epidemiology); epidemiologic methods; HIV; influenza; vaccines; validation sets

Protective vaccine efficacy for susceptibility, VE_S, is usually measured as $VE_S = 1 - RR$, where RR is some measure of relative risk in the vaccinated group compared with the unvaccinated group (1, 2). In many studies of vaccines, such as vaccines for influenza, rotavirus, pertussis, and cholera, confirmatory diagnosis of a suspected case is done by culture or a quick test of a swab, sputum, blood, or stool sample. If all ascertained cases were actually cases of the disease of interest and most of the cases were ascertained, the estimates would be accurate.

However, samples or cultures are often expensive or difficult to collect, so a less specific case definition is used. In an influenza study, a nonspecific case definition might be "any respiratory illness" (3) or "febrile upper respiratory tract illness" (4). Thus, many ascertained cases are not cases of the disease for which vaccination confers protection. This severely attenuates efficacy estimates. For instance, using only culture-confirmed cases, Belshe et al. (5) estimated the protective efficacy of a live attenuated influenza vaccine in a randomized controlled trial in children to be 0.89 (also see Belshe et al. (6) and Longini et al. (7)). Using a case definition of "upper respiratory tract illness with either fever or cough," Nichol et al. (4) estimated the protective efficacy of a similar live attenuated influenza vaccine in a randomized controlled trial in adults to be only 0.25.

Indirect and overall effectiveness measures are obtained by comparing the risk of disease in a community that has a vaccination program with the risk in a community that has no vaccination or has a different program. In a group-randomized influenza vaccine trial for overall effectiveness (3), 30 schools were randomly assigned to receive either vaccine or placebo. The outcome was the presence of one or more respiratory illnesses or the absence of such illnesses during the epidemic

Received for publication June 13, 2000, and accepted for publication October 26, 2000.

Abbreviations: IP, incidence proportion; VE, vaccine efficacy.

From the Department of Biostatistics, Rollins School of Public Health, Emory University, 1518 Clifton Road NE, Atlanta, GA 30322. (Reprint requests to Dr. M. Elizabeth Halloran at this address (e-mail: mehallo@sph.emory.edu)).

period. The low estimate of overall effectiveness of a live candidate vaccine (e.g., 1 – [430/2,525]/[567/2,331] = 0.30 in one age group) could easily have resulted from the nonspecific case definition. If just half of the respiratory illness in the placebo group were not influenza, corresponding to 307 noninfluenza cases in the vaccine group, the effectiveness point estimate would be 0.59 (1 – [127/2,525]/[283/2,331] = 0.59)—a factor of 2 greater.

The efficacy of a vaccine in reducing infectiousness, VE_I, is usually measured by comparing the risk of transmission from a vaccinated person with the risk of transmission from an unvaccinated person (2). However, data on exposure to infection are often difficult or expensive to obtain and are inherently prone to mismeasurement.

Study designs that combine data of different levels and quality have been developed in some areas of epidemiology, such as nutritional epidemiology and cancer epidemiology. In a small subset of participants, called the *validation sample*, a good measurement of the exposure or outcome of interest is obtained. For each of the participants, including the validation set, a less accurate or coarser, possibly cheaper, measure is also obtained. The less accurate measure is sometimes called a "surrogate." The better measure is sometimes called the "good data" or even the "gold standard." Estimates based on just the surrogate measure would usually be biased. Estimates based on just the good data from the small validation sample would not be very precise. The general idea is that the good data in the validation set correct the bias, while the larger main study increases precision.

In a previous article (8), we suggested that these methods could be extended to improve the design and analysis of studies of infectious disease, particularly in vaccine evaluation. In this paper, we show how studies with validation sets can produce more accurate and efficient estimates in vaccine field studies. We discuss challenges posed by infectious diseases in the use of currently available methods and call for more methodological developments.

STUDIES WITH VALIDATION SETS

One approach to the validation method is to have a validation sample within a cohort study. A related approach includes *two-stage* case-control studies. At the first stage, information is obtained on the outcome and on a subset of the covariates. At the second stage, more accurate information might be obtained about the outcome or the already-measured covariates, or additional covariates might be measured in a subsample of the cases and controls (9–14).

Many statistical methods are available for analyzing studies with validation sets (12, 15–23). Usually one starts with a parametric model, such as a likelihood model or an estimating equation, for the people on whom good data are available. Then one uses some method to combine the people on whom coarser data are available into the analysis. Full likelihood or Bayesian approaches usually model the relation of the coarser data to the more accurately measured data and build the model into the analysis. However, the relation between the good measure and the less accurate measure is not of any scientific interest. If the model relating the two measures is wrong, the analysis can be quite biased.

Semiparametric methods either use a nonparametric method to estimate the relation or do not estimate it all. By avoiding parametric specification of the relation between the good covariate and the surrogate, semiparametric methods avoid the bias that results from misspecifying the relation. An example is the semiparametric mean score method for outcomes (24) and for covariates (23). In the case of covariates or exposure variables, the *score* contribution (i.e., the derivative of the log likelihood) for each main study member on whom only coarse exposure data are available is estimated from the average score contributions of the validation sample members with the same observed covariate and outcome values. The mean score approach for outcomes is similar in that a surrogate outcome is measured in everyone, while the accurate outcome is measured only in the validation sample. The semiparametric efficient methods of Robins et al. (20, 21) also avoid nonparametric estimation of the missing covariate distribution. However, the semiparametric efficient methods extract further information from people who are not in the validation set.

Several approaches have been developed for using validation sets for outcome data (15, 25–27). In the survival setting, much work related to missing or mismeasured covariates has been conducted (28–31), but less work has been done for misclassified outcomes.

USING VALIDATION SETS FOR OUTCOMES

Consider estimating protective efficacy for susceptibility, VE_S, in a randomized, double-blinded, placebo-controlled trial of an influenza vaccine in children. Suppose that we want to estimate VE_S based on the relative incidence proportion (IP), also known as the attack rate or cumulative incidence, at the conclusion of the study. Assume that through our definition of an influenza-like illness, we ascertain every occurrence of true influenza illness in the study population. We then confirm each suspected illness by performing a culture for influenza. Assume that the sensitivity and specificity of a culture are both 100 percent. Let N_1 and N_0 be the numbers of children in the vaccinated and unvaccinated groups, respectively. Let y_1 and y_0 be the numbers of influenza cases in the vaccinated and unvaccinated groups, respectively. Then the efficacy estimate is

$$VE_{S,IP} = 1 - \frac{IP_1}{IP_0} = 1 - \frac{y_1/N_1}{y_0/N_0},$$

where IP_1 and IP_0 are the incidence proportions in the vaccinated and unvaccinated groups, respectively.

Suppose, however, that we do not confirm suspected cases by culture. The ascertained cases of influenza-like illness will possibly include many cases that are not influenza but illnesses caused by other viruses, such as respiratory syncytial virus or parainfluenza. We will call such illnesses

"noninfluenza." The term "influenza-like illness" captures both the true influenza cases and the noninfluenza cases. Exactly what the terms include will depend on the definitions used in any particular study.

Suppose that z_1 and z_0 are the numbers of noninfluenza cases in the vaccinated and unvaccinated groups, respectively. Then the total number of influenza-like illnesses in vaccine group v, $v = 0, 1$, is $w_v = z_v + y_v$. The efficacy estimate based on the total number of influenza-like illnesses would be

$$\mathrm{VE}_{S,\mathrm{IP},a} = 1 - \frac{\mathrm{IP}_{1,a}}{\mathrm{IP}_{0,a}} = 1 - \frac{w_1/N_1}{w_0/N_0},$$

where a denotes all influenza-like illness.

Consider the example shown in table 1. Our estimate of VE_S based on the true influenza cases would be

$$\mathrm{VE}_{S,\mathrm{IP}} = 1 - \frac{10/1{,}000}{100/1{,}000} = 0.90 \text{ (95 percent confidence interval: 0.81, 0.95).}$$

Our estimate based on all-influenza like illness would be

$$\mathrm{VE}_{S,\mathrm{IP},a} = 1 - \frac{60/1{,}000}{150/1{,}000} = 0.60 \text{ (95 percent confidence interval: 0.45, 0.71).}$$

A simple adjustment using a validation set

How could a validation sample help correct the attenuated VE_S estimate obtained using all influenza-like illnesses? Continuing the above example, assume that the incidence rates of both influenza and noninfluenza are constant over time. We randomly sample a fraction of the vaccinated and the unvaccinated influenza-like cases, and we culture swabs taken from them to confirm whether they had true influenza. From the results, we can estimate the probability in each group that any influenza-like case is true influenza.

We denote the sampling fraction as ρ_v, the number of influenza-like cases sampled for the validation set as r_v, and the number of culture-confirmed influenza cases as c_v, $v = 0, 1$. We estimate the proportion π_v of the influenza-like cases that are true influenza from the ratio of the number of culture-confirmed cases to the total number of influenza-like cases in each validation set, i.e., $\hat{\pi}_v = c_v/r_v$, $v = 0, 1$. This estimated proportion is used to adjust the number of influenza-like illnesses in each vaccine arm to estimate the number of true influenza cases.

TABLE 1. Results of a hypothetical influenza vaccine trial in children

Exposure	Cases			No. of children (N_v)
	Influenza (y_v)	Non-influenza (z_v)	All influenza-like ($w_v = y_v + z_v$)	
Vaccinated ($v = 1$)	10	50	60	1,000
Unvaccinated ($v = 0$)	100	50	150	1,000

Suppose that the sampling fractions are $\rho_1 = 0.20$ and $\rho_0 = 0.10$ in the vaccinated and unvaccinated groups, respectively. Then we would expect to sample $r_1 = 0.20(60) = 12$ and $r_0 = 0.10(150) = 15$ influenza-like cases for the vaccinated and unvaccinated validation samples, respectively. We would expect 10/60 of the cultured vaccinated cases to be true influenza, or $c_1 = 2$ of the $r_1 = 12$ cases in the validation sample. We estimate $\hat{\pi}_1 = 2/12 = 0.17$. Similarly, we would expect 100/150 of the cultured unvaccinated influenza-like cases to be culture-confirmed influenza—that is, $c_0 = 10$ of the $r_0 = 15$ cases in the validation sample. We estimate $\hat{\pi}_0 = 10/15 = 0.67$. We then multiply the observed number of all influenza-like illnesses in each group by the estimated proportion of true influenza to obtain

$$\widehat{\mathrm{VE}}_{S,\mathrm{IP},v} = 1 - \frac{(\hat{\pi}_1 w_1)/N_1}{(\hat{\pi}_0 w_0)/N_0} \quad (1)$$

$$= 1 - \frac{[0.17(60)]/1{,}000}{[0.67(150)]/1{,}000}$$

$$= 0.90 \text{ (95 percent confidence interval: 0.64, 0.98),}$$

where the subscript v denotes validation set. The approximate 95 percent confidence interval is based on the point estimate from the adjusted incidence proportions plus or minus 1.96 times the standard error obtained using the mean score method (24) with a logistic model (odds ratio). In this case, we are using the standard error of an odds ratio as an approximation for the standard error of a risk ratio. For consistency of comparison, the 95 percent confidence intervals reported above for vaccine efficacy based on the full data and all influenza-like illness were also calculated using a logistic model, while calculating the point estimate using the relative incidence proportions. In general, approximate standard errors based on the risk ratios or odds ratios are very similar. In this case, the confidence intervals based on the two approaches differed by less than 0.01.

The simple adjustment corrects for the bias resulting from using influenza-like illness as the outcome without our having to culture every suspected case. The main penalty in using the validation sample rather than culturing everyone is the increased uncertainty in the estimate. The variability of the estimate obtained using a validation sample depends on the size of the validation set. In this example, if the sampling fraction in each group were doubled, the approximate 95 percent confidence interval would decrease to (0.74, 0.96).

The degree of attenuation of the VE_S estimates from using the nonspecific case definition depends on the ratio of true disease to background nonspecific disease. In the above example, if instead of 50 there had been 100 noninfluenza

cases in each group, the estimate based on all influenza-like illness would have been

$$\widehat{\mathrm{VE}}_{S,\mathrm{IP},a} = 1 - \frac{(10 + 100)/1{,}000}{(100 + 100)/1{,}000} = 0.45,$$

an even worse attenuation. However, validation sets would still be able to adjust this estimate. If a vaccine is highly efficacious, it might be desirable to have a higher sampling fraction in the vaccinated group. In general, each stratum of interest could have a different sampling fraction.

Time-varying incidence rates

The incidence of true influenza, as well as the incidence of other noninfluenza illnesses, varies rapidly during a typical influenza season. Thus, the ratio of influenza cases to noninfluenza cases can vary greatly during a study. Any method of adjustment will need to take this time variability into account.

For example, suppose that we group the influenza-like cases within small time intervals τ, such as 1 week: $(t_{\tau-1}, t_\tau]$, $\tau = 1, \ldots, T$. If the influenza epidemic or vaccine study lasts 12 weeks, then $T = 12$. We also group the validation samples $r_{v\tau}$ within time intervals. Then we estimate the proportion $\pi_{v\tau}$ of true influenza cases among the influenza-like illnesses in each vaccine group v, $v = 0, 1$, from the validation samples in each time interval τ, $\tau = 1, \ldots, T$; that is, $\hat{\pi}_{v\tau} = c_{v\tau}/r_{v\tau}$. We multiply the number of influenza-like illnesses ascertained in each week $w_{v\tau}$ by the estimated $\{\pi_{v\tau}\}$ for that time interval to obtain an adjusted estimate of the number of influenza cases in each interval. Summing over the adjusted estimates of the number of true influenza cases in each interval, we obtain an adjusted estimate of the total number of influenza cases in each group during the study. From this, we estimate the incidence proportion of true influenza in each vaccine group, and from that, $\mathrm{VE}_{S,\mathrm{IP}v}$:

$$\widehat{\mathrm{VE}}_{S,\mathrm{IP},v} = 1 - \frac{[\Sigma_{\tau=1}^{T}\hat{\pi}_{1\tau}w_{1\tau}]/N_1}{[\Sigma_{\tau=1}^{T}\hat{\pi}_{0\tau}w_{0\tau}]/N_0}. \qquad (2)$$

Smoothing methods (32) could be used on the $\{\pi_{v\tau}\}$.

Figure 1 shows the results of 100 simulations for estimating vaccine efficacy based on 1) true influenza, 2) the use of this simple validation set approach, and 3) all influenza-like illness. The influenza epidemic in this example lasted for 12 weeks. The expected incidence in children varied weekly as (0.014, 0.024, 0.034, 0.05, 0.06, 0.055, 0.05, 0.044, 0.038, 0.024, 0.015, 0.01). The expected incidence rate of noninfluenza was set to 0.02 per week. The expected weekly incidence rates of influenza and noninfluenza were each multiplied by an independent uniform random number between 0.85 and 1.15. Since both noninfluenza incidence and true influenza incidence were multiplied by random numbers, the ratio of true influenza to noninfluenza varied among simulations. The set vaccine efficacy was $\mathrm{VE}_S = 0.90$ with a multiplicative (leaky) effect. In each week, we sampled $\rho_0 = 0.25$ and $\rho_1 = 0.40$ of the influenza-like illnesses in the unvaccinated children and the vaccinated children, respectively.

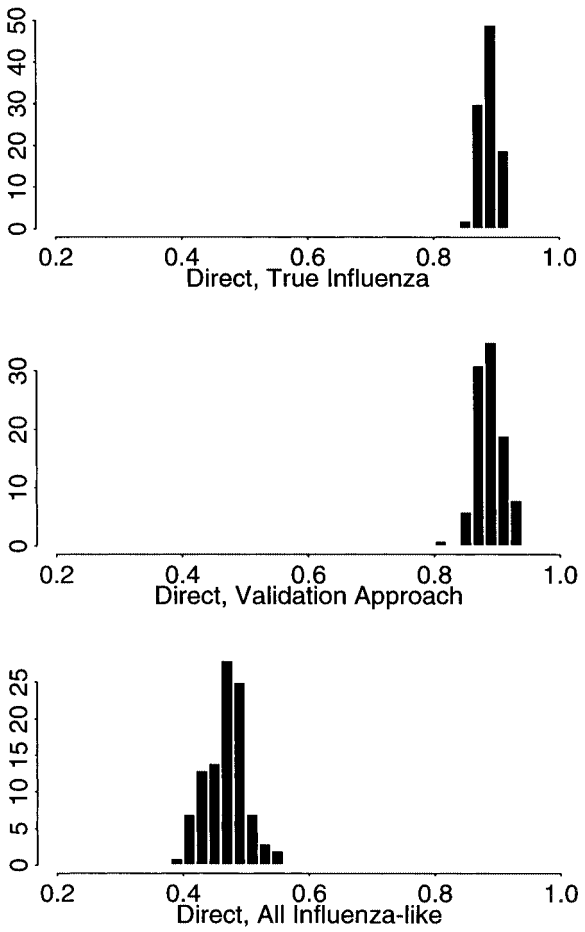

FIGURE 1. Vaccine efficacy for susceptibility in children as estimated using true influenza cases (top), the validation set approach (middle), and all influenza-like illness (bottom). Results were based on 100 simulations. Vaccine efficacy was set to 0.90 using a multiplicative (leaky) model. Other simulation parameters are described in the text.

In figure 1, the $\mathrm{VE}_{S,\mathrm{IP}}$ estimates based on true influenza cases vary between 0.85 and 0.95, with a median at 0.89, just below the simulated value of 0.90. However, the VE_S estimate based on all influenza-like illness varies between about 0.40 and 0.60, with a median at 0.47. With the use of the simple adjustment based on the validation set, the estimates are again centered around 0.89, though the variability is greater than it would have been if all true influenza cases had been ascertained.

Validation sets in community trials

The usefulness of validation sets for outcomes may be even greater in community trials designed to estimate the indirect and overall effects of vaccination programs. In large community trials, culturing every suspected case may be prohibitively expensive, as well as operationally unfeasible. The purpose of the studies would primarily be to estimate

the indirect effects of vaccination on people who were not vaccinated and the overall benefit to the community of widespread vaccination (2, 33, 34). The denominators for the estimation could be either the number of people in each relevant stratum in the community as a whole, a health maintenance organization catchment population, or some other relevant catchment population for the observed cases.

Many features complicate community-based vaccination studies. Chief among them is the comparability of the communities included in the study with respect to the baseline incidence and the background incidence of any disease included in a nonspecific case definition. Even if the communities are comparable, however, a nonspecific case definition can attenuate the estimates of indirect and overall effects.

In figures 2 and 3, we present results of 100 simulated estimates of the indirect effects of vaccinating 50 percent of the children in one community as compared with another community without vaccination. This scenario is similar to that depicted in figure 1, with 10,000 people in each population, half children and half adults. The incidence rate of true influenza in adults is only half that in children, while the incidence rate of noninfluenza in adults is the same as that in children. The baseline incidences of true influenza and background noninfluenza are multiplied by random numbers between 0.85 and 1.15, so the baseline incidences in the two comparison communities are similar but not identical. To estimate indirect effects in children (figure 2), one compares the incidence proportion among unvaccinated children in the community that has the vaccination program with the incidence proportion among (unvaccinated) children in the community without vaccination. A similar comparison is made among the adults (figure 3), all of whom are unvaccinated. We have set the indirect effects to 0.25. The top histograms of estimates based on ascertainment of all true influenza cases in children and adults are centered around 0.25, the set value. However, if we use all influenza-like illnesses, the estimates are much lower (bottom rows). The histogram is centered around 0.14 in children and 0.10

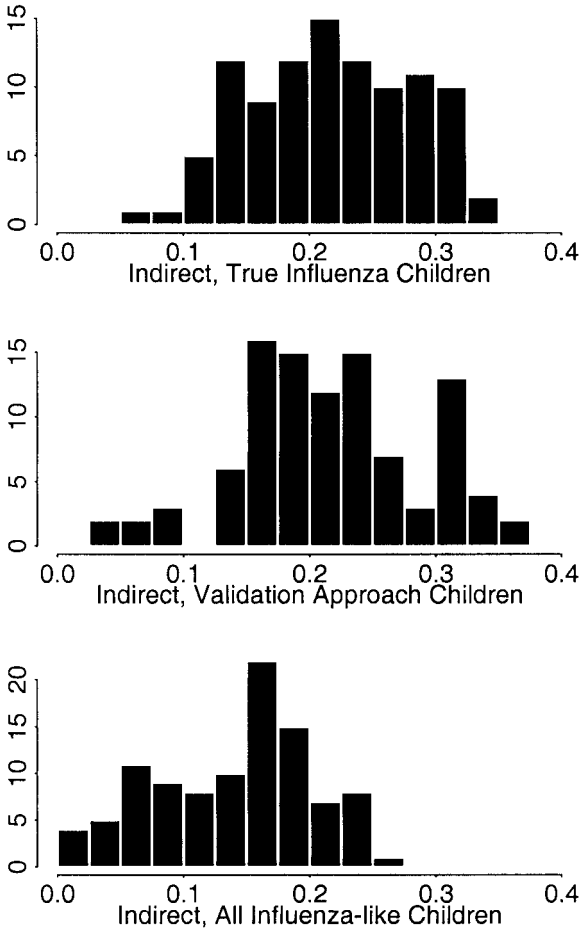

FIGURE 2. Estimated indirect effects of vaccination of children among children in a community trial when the indirect effects are set to 0.25. Estimates were based on true influenza cases (top), the validation set approach (middle), and all influenza-like illness (bottom), from 100 simulations. Other simulation parameters are described in the text.

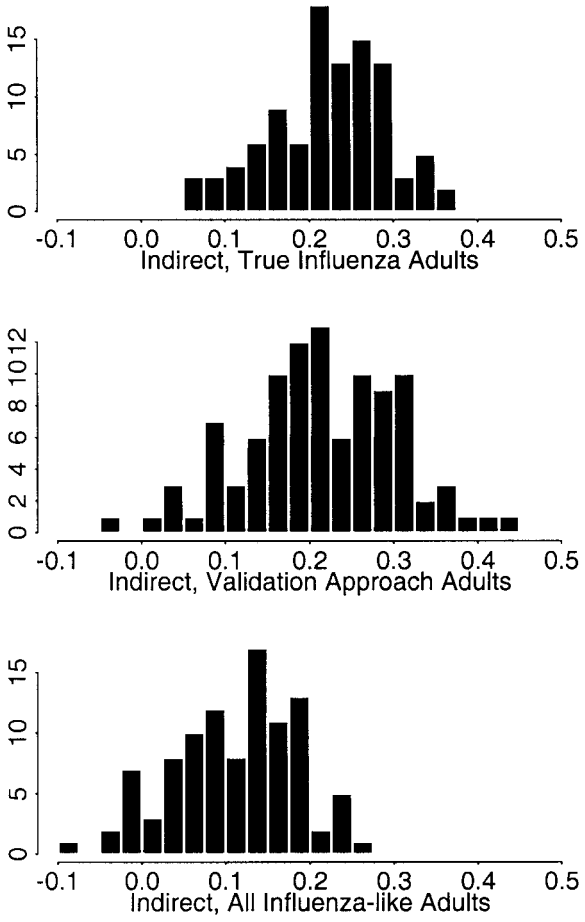

FIGURE 3. Estimated indirect effects of vaccination of children among adults in a community trial when the indirect effects are set to 0.25. Estimates were based on true influenza cases (top), the validation set approach (middle), and all influenza-like illness (bottom), from 100 simulations. Other simulation parameters are described in the text.

in adults. Using the simple time-varying adjustment with validation sets that was described above (see previous section), the adjusted indirect effect estimates are once again centered more closely around 0.25.

Validation sets as discussed here do not adjust for noncomparability between communities. However, as figures 2 and 3 illustrate, validation sets can help investigators retrieve a better estimate of indirect effects than is obtained by using a nonspecific case definition. For example, in an influenza vaccine study in Texas designed to evaluate the indirect effects on adults of vaccinating children (35), the primary outcome was medically attended acute respiratory illness during the influenza season. Use of a validation set could help investigators obtain better estimates of the indirect effects on influenza in adults, even if the comparison communities were not completely comparable.

USING VALIDATION SETS FOR EXPOSURE TO INFECTION

In simulation studies, Golm et al. (36, 37) explored using validation samples for exposure-to-infection information to estimate vaccine efficacy for infectiousness, VE_I, in trials of human immunodeficiency virus vaccine. The complete data in such a trial are based on the augmented vaccine trial design (38–40). The idea is to recruit and randomize individuals to the vaccine trial but also try to recruit steady partners of some of the primary participants. The complete exposure data in the trial would include the number of sexual contacts of the primary participants and recruited partners with people outside the partnership, the number of sexual contacts within each partnership, and the infection status of the partners. In this setting, the relative risk estimate for VE_I is based on the relative per-contact transmission probability within each partnership. However, accurate information on sexual contacts is difficult to collect.

With a focus on the potential for improving estimation of VE_I, Golm et al. (37) assumed that only partnerships were included in the validation sample. In the validation partnerships, information on sexual contacts was assumed to be gathered without error. Thus, an easy, coarse measure of exposure consisted of each partnership's classifying their number of within-partnership contacts as either high or low (Hi/Lo). Semiparametric analytical methods (22, 23) were then applied. A surrogate M was also assumed as measured from each partnership's making some guess at their number of sexual contacts.

Figure 4 illustrates the potential for improving estimates of VE_I. The histograms shown are from 200 simulations of a human immunodeficiency virus vaccine trial with 4,000 primary trial participants and 2,000 with steady partners. The sampling fraction was 0.20 of the partnerships for the validation set. Vaccine efficacy was set to $VE_S = 0.4$ and $VE_I = 0.6$. (For more details, see Golm et al. (37).) The top histogram presents estimates based on participants and partners for whom complete, good data are available (complete cases). The estimates for VE_I are quite variable, since there is little information available. In the next two histograms, the two semiparametric approaches provide much more precise estimates than the complete case estimates. These methods incorporate the information from the main study on people for whom only coarse exposure data are available. The surrogate M (fourth histogram) actually performs fairly well. However, using the coarse data based on Hi/Lo alone yields a very biased, though precise, estimate of VE_I (bottom histogram). The problem of bias is overcome with the use of the validation set.

DISCUSSION

Validation sets for outcomes and for exposure to infection have much potential for improving the precision and accuracy of estimates from field studies of infectious diseases, espe-

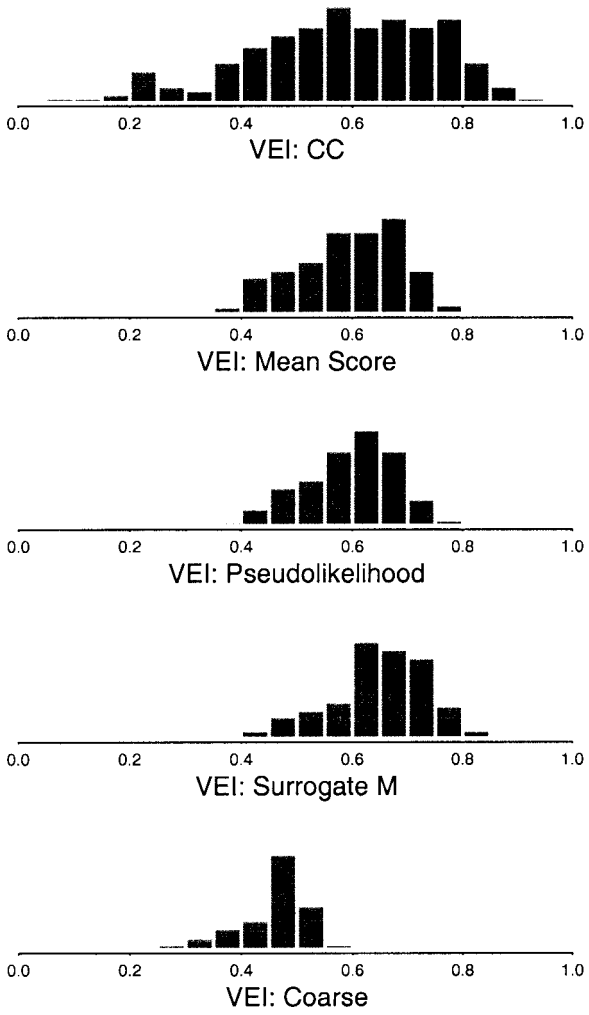

FIGURE 4. Histograms of Monte Carlo estimates of vaccine efficacy for infectiousness (VEI), using five different methods of estimation (see text). The histograms shown are from 200 simulations of a human immunodeficiency virus vaccine trial with 4,000 primary trial participants and 2,000 with steady partners (37). The sampling fraction was 0.20 of the partnerships for the validation set. The value of VEI used for simulation was 0.6. CC, complete cases; surrogate *M*, surrogate measure obtained from each partnership's guess at their number of sexual contacts. (Adapted from Golm et al. (37).)

cially in evaluating vaccines and vaccination strategies. For prelicensure primary efficacy trials, it is less likely that such methods will replace the current practice of using confirmatory laboratory diagnosis. However, in secondary or postlicensure studies or in large community-level studies, these methods could be widely applicable and cost-effective.

Several challenges to use of validation sets are posed by the infectious disease setting. In the simple, time-invariant influenza example presented above, existing methods could be applicable. However, the rapidly time-varying incidence rates of some infectious diseases present new problems. The probability that any suspected case is a case of the disease of interest changes rapidly over time. Methods employed must take this rapid time evolution into account. A person might have more than one event of misclassified disease during a study (41–43), but generally a person would have only one case of the disease of interest. This raises issues related to the validation sampling scheme. The problem with sampling individuals when they sometimes present with influenza-like illness and sometimes do not is that it is possible to miss sampling them when they have true influenza. Presumably, then, when they presented again, they actually would no longer be in the risk set for having influenza. This problem could be avoided by selecting people for the validation set before the study begins and culturing them each time they present with illness fitting the nonspecific case definition. Such issues require further examination.

Other concerns in selection of the validation sample are common to the study of noninfectious diseases. The validation set may not be internal to the actual study but may be drawn from some other external population. For example, influenza epidemics are often monitored by culturing people with suspected influenza cases once the season has begun. The samples so cultured are usually convenience samples. Whether a physician decides to take a culture can be heavily influenced by his or her belief as to whether the person has true influenza. If such convenience samples were used uncritically to adjust vaccine efficacy or effectiveness estimates, the results could be very biased. However, methods for using such convenience samples could be developed. If a physician knows a person's vaccine status, it might affect whether he or she takes a sample. People who are vaccinated may have less serious illness and may tend not to visit a physician or report symptoms. Optimal sampling strategies also have yet to be explored in this context. Efficient methods will probably vary the probability of being selected into the validation set with time, as well.

Other problems, such as whether the good measure of outcome is a gold standard or itself is prone to mismeasurement, need further examination. The probability of obtaining a positive culture may depend on the vaccination status of an individual, because vaccination could shorten the period in which a positive culture can be obtained or could reduce the shedding of the infectious agent so that the culture is less likely to be positive. The choice of the nonspecific case definition is also important in determining the ratio of true cases of interest to background cases. Similar problems in using validation sets for exposure to infection remain to be solved. Of particular importance is the fact that there is really no gold standard for exposure to infection in any setting. These and many other issues could be fruitfully examined to improve vaccine efficacy and effectiveness studies.

Analytical methods for combining participants with different levels of data on exposure to infection could also be used in new approaches for estimating VE_S. In vaccine trials, the primary analysis generally uses one of the unconditional estimates of VE_S. Often information on contact with and exposure to infection is available on some of the participants, more by happenstance than by design. An example might be a pertussis vaccine trial in which information on household exposure to infection is available for some of the participants but no exposure is observed for most participants. Until now, the subset of individuals for whom exposure information was available was analyzed in a second analysis to obtain VE_S and, less often, VE_I estimates based on the transmission probability or secondary attack rates. However, methods could be developed to incorporate the different levels of information into a single analysis to improve estimation of VE_S as well as VE_I.

In this paper, we have considered separately the use of validation sets for outcomes and for exposure to infection. However, it would be possible for studies to use validation sets for both. Validation sets could be used in other infectious disease studies as well. In malaria studies, exposure to infection is measured by capturing mosquitoes. Validation sets for exposure to malaria could be useful in studies on developing immunity. The potential for using validation sets in infectious disease field studies has just begun to be explored. There is room for many new developments to meet the special challenges of studying infectious diseases.

ACKNOWLEDGMENTS

This research was partially support by grants R01-A132042 and R01-A140846 from the National Institutes of Health.

REFERENCES

1. Halloran ME, Longini IM Jr, Struchiner CJ. Design and interpretation of vaccine field studies. Epidemiol Rev 1999;21:73–88.
2. Halloran ME, Struchiner CJ, Longini IM Jr. Study designs for different efficacy and effectiveness aspects of vaccination. Am J Epidemiol 1997;146:789–803.
3. Rudenko LG, Slepushikin AN, Monto AS, et al. Efficacy of live attenuated and inactivated influenza vaccines in schoolchildren and their unvaccinated contacts in Novgorod, Russia. J Infect Dis 1993;168:881–7.
4. Nichol KL, Mendelman PM, Mallon KP, et al. Effectiveness of live, attenuated intranasal influenza virus vaccine in healthy working adults: a randomized controlled trial. JAMA 1999;282:137–44.
5. Belshe RB, Mendelman PM, Treanor J, et al. The efficacy of live attenuated, cold-adapted, trivalent, intranasal influenza-virus vaccine in children. N Engl J Med 1998;20:1405–12.
6. Belshe RB, Gruber WC, Mendelman PM, et al. Efficacy of vaccination with live attenuated, cold-adapted, trivalent,

intranasal influenza virus vaccine against a variant (A/Sydney) not contained in the vaccine. J Pediatrics 2000;136:168–75.
7. Longini IM Jr, Halloran ME, Nizam A, et al. Estimation of the efficacy of live, attenuated influenza vaccine from a two-year, multi-center vaccine trial: implications for influenza epidemic control. Vaccine 2000;18:1902–9.
8. Halloran ME, Longini IM Jr, Struchiner CJ, et al. Exposure efficacy and change in contact rates in evaluating prophylactic HIV vaccines in the field. Stat Med 1994;13:357–77.
9. White JE. A two stage design for the study of the relationship between a rare exposure and a rare disease. Am J Epidemiol 1982;115:119–28.
10. Cain KC, Breslow NE. Logistic regression analysis and efficient design for two-stage studies. Am J Epidemiol 1988;128: 1198–206.
11. Breslow NE, Cain KC. Logistic regression for two-stage case-control data. Biometrika 1988;75:11–20.
12. Flanders WD, Greenland S. Analytic methods for two-stage case-control studies and other stratified designs. Stat Med 1991;10:739–47.
13. Zhao LP, Lipsitz S. Designs and analysis of two-stage studies. Stat Med 1992;11:769–82.
14. Breslow NE, Holubkov R. Maximum likelihood estimation of logistic regression parameters under two-phase, outcome dependent sampling. J R Stat Soc B 1997;59:447–61.
15. Pepe MS. Inference using surrogate outcome data and a validation sample. Biometrika 1992;79:355–65.
16. Carroll RJ, Wand WP. Semiparametric estimation in logistic measurement error models. J R Stat Soc B 1991;53:573–85.
17. Carroll RJ, Ruppert D, Stefanski LA. Measurement error in nonlinear models. London, United Kingdom: Chapman and Hall Ltd, 1995.
18. Little RJ, Rubin DB. Statistical analysis with missing data. New York, NY: John Wiley and Sons, Inc, 1987.
19. Rubin DB. Inference and missing data. Biometrika 1976;63: 581–92.
20. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. J Am Stat Assoc 1994;89:846–66.
21. Robins JM, Hsieh F, Newey W. Semiparametric efficient estimation of a conditional density with missing or mismeasured covariates. J R Stat Soc B 1995;57:409–24.
22. Pepe MS, Fleming TR. A nonparametric method for dealing with mismeasured covariate data. J Am Stat Assoc 1991;86:108–13.
23. Reilly M, Pepe MS. A mean score method for missing and auxiliary covariate data in regression models. Biometrika 1995; 82:299–314.
24. Pepe MS, Reilly M, Fleming TR. Auxiliary outcome data and the mean score method. J Stat Plan Inference 1994;42:137–60.
25. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med 1989;8:431–40.
26. Wittes J, Lakatos E, Probsfield J. Surrogate endpoints in clinical trials: cardiovascular diseases. Stat Med 1989;8:415–25.
27. Magder LS, Hughes JP. Logistic regression when the outcome is measured with uncertainty. Am J Epidemiol 1997;146:195–203.
28. Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. Biometrika 1982;69:331–42.
29. Robins JM, Rotnitzky A. Recovery of information and adjustment for dependent censoring using surrogate markers. In: Jewell N, Dietz K, Farewell V, eds. AIDS epidemiology—methodological issues. Cambridge, MA: Birkhäuser Boston, 1992:297–331.
30. Pugh M, Robins JM, Lipsitz S, et al. Inference in the Cox proportional hazards model with missing covariates. (Technical report). Boston, MA: Department of Biostatistics, Harvard School of Public Health, 1993.
31. Tsiatis AA, DeGruttola V, Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error: applications to survival and CD4 counts in patients with AIDS. J Am Stat Assoc 1995;90:27–37.
32. Hastie TJ, Tibshirani RJ. Generalized additive models. London, United Kingdom: Chapman and Hall Ltd, 1990.
33. Halloran ME, Struchiner CJ. Study designs for dependent happenings. Epidemiology 1991;2:331–8.
34. Hayes RJ, Alexander ND, Bennett S, et al. Design and analysis issues in cluster-randomized trials of interventions against infectious diseases. Stat Methods Med Res 2000;9:95–116.
35. World Health Organization. Influenza vaccine. Wkly Epidemiol Rec 2000;75:281–8.
36. Golm GT, Halloran ME, Longini IM Jr. Semiparametric models for mismeasured exposure information in vaccine trials. Stat Med 1998;17:2335–52.
37. Golm GT, Halloran ME, Longini IM Jr. Semiparametric methods for multiple exposure mismeasurement and a bivariate outcome in HIV vaccine trials. Biometrics 1999;55:94–101.
38. Longini IM Jr, Datta S, Halloran ME. Measuring vaccine efficacy for both susceptibility to infection and reduction in infectiousness for prophylactic HIV-1 vaccines. J Acquir Immune Defic Syndr Hum Retrovirol 1996;13:440–7.
39. Datta S, Halloran ME, Longini IM Jr. Augmented HIV vaccine trial designs for estimating reduction in infectiousness and protective efficacy. Stat Med 1998;17:185–200.
40. Longini IM Jr, Hudgens MG, Halloran ME, et al. A Markov model for measuring vaccine efficacy for both susceptibility to infection and reduction in infectiousness for prophylactic HIV-1 vaccines. Stat Med 1999;18:53–68.
41. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes under the presence of missing data. J Am Stat Assoc 1995;90:106–21.
42. Rotnitzky A, Robins JM. Semiparametric regression estimation in the presence of dependent censoring. Biometrika 1995;82: 805–20.
43. Rotnitzky A, Robins JM, Scharfstein D. Semiparametric regression for repeated outcomes with non-ignorable nonresponse. J Am Stat Assoc 1998;444:1321–40.